

基于 PSO-SVM-RF 的离心泵转子故障诊断研究

范传翰¹, 宋礼威², 刘厚林¹, 董 亮¹, 肖幸鑫¹, 陈泽宇¹

(1. 江苏大学流体机械工程技术研究中心, 江苏 镇江 212000;

2. 中广核工程有限公司核电安全监控技术与装备国家重点实验室, 广东 深圳 518172)

摘 要: 转子是离心泵的重要组件之一, 对转子运行状态的检测和诊断的研究具有重要意义。转子不平衡、不对中故障引发的故障特征较为相似, 为了有效识别离心泵这两种转子故障。通过在离心泵进口法兰位置布置振动加速度传感器进行信号采集, 提取原信号时频域特征, 并利用随机森林算法筛选出重要性较高6个特征并将随机森林得到的分类结果作为 PSO-SVM 的输入, 进而来区分正常、转子不对中、转子不平衡故障, 同时还比较了该方法与传统 PSO-SVM 的故障识别率。结果表明, 该模型 PSO 优化迭代次数更少、具有更高的识别率, 对故障的识别率达到 99.36%。

关键词: 离心泵; 故障诊断; 支持向量机; 随机森林; 转子不平衡; 转子不对中

中图分类号: TH311

文献标识码: A

DOI: 10.12396/znsd.220656

范传翰, 宋礼威, 刘厚林, 等. 基于 PSO-SVM-RF 的离心泵转子故障诊断研究[J]. 中国农村水利水电, 2023(2): 171-176. DOI: 10.12396/znsd.220656.

FAN C H, SONG L W, LIU H L, et al. Research on the rotor fault diagnosis of the centrifugal pump based on PSO-SVM-RF [J]. China Rural Water and Hydropower, 2023(2): 171-176. DOI: 10.12396/znsd.220656.

Research on the Rotor Fault Diagnosis of the Centrifugal Pump Based on PSO-SVM-RF

FAN Chuan-han¹, SONG Li-wei², LIU Hou-lin¹, DONG Liang¹, XIAO Xing-xin¹, CHENG Ze-yu¹

(1. Fluid Machinery Engineering Technology Research Center of Jiangsu University, Zhenjiang 212000, Jiangsu Province, China;

2. State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment of CGN Engineering Co., Ltd, Shenzhen 518172, Guangdong Province, China)

Abstract: Rotor is one of the important components of centrifugal pump, which is of great significance to the detection and diagnosis of rotor running state. The characteristics of faults caused by unbalanced and misaligned rotor faults are similar. In order to effectively identify the two rotor faults of centrifugal pump rotation, the influence of location selection of different measuring points on distinguishing faults is effectively understood. Sensors are arranged on the inlet flange of centrifugal pump for signal acquisition, and the time-frequency domain characteristics of the original signal are extracted. The random forest algorithm is used to screen out 6 features with high importance, and the classification results obtained by the random forest are used as the input of PSO-SVM, so as to distinguish the faults of normal, misaligned rotor and unbalanced rotor. At the same time, the fault recognition rate of this method is compared with that of traditional PSO-SVM. The results show that the PSO model has fewer iterations and higher recognition rate, and the recognition rate of faults reaches 99.36%.

Key words: centrifugal pump; fault diagnosis; support vector machine; random forests; imbalance of rotary; misalignment of rotary

收稿日期: 2022-04-06

基金项目: 国家自然科学基金项目(51879122, 51779108, 51779106); 镇江市重点研发计划项目(GY2017001, GY2018025); 西华大学流体及动力机械教育部重点实验室开放课题项目(szjj2017-094, szjj2016-068); 江苏高校优势学科建设工程项目(PAPD); 江苏省“六大人才高峰”高层次人才项目(GBZB-017)。

作者简介: 范传翰(1993-), 男, 主要研究方向为离心泵机组故障诊断。E-mail: 79675590@qq.com。

通讯作者: 董 亮(1981-), 男, 研究员, 博士生导师, 主要研究方向为泵监测及故障诊断。E-mail: dongliang@ujs.edu.cn。

0 引言

离心泵广泛应用于国民经济的各个领域。由于离心泵工作环境复杂,运行工况调节频繁从而导致离心泵发生故障的概率增大,造成效率降低,甚至会出现安全事故。转子是离心泵中的重要组件之一,机组在长期运行过程中容易导致转子发生故障。其中,转子不平衡、不对中故障最为常见^[1]。近年来,随着计算机技术的进步,国内外学者利用不同的方法为不同设备的故障诊断提供了很多参考^[2]。屈梁生^[3]系统研究了全息谱技术所衍生各项技术及在各类机械故障诊断实践中的应用。Dong^[4]等人将小波包分解(WPD)与主成分分析(PCA)和径向基函数(RBF)神经网络相结合,完成了对离心泵汽蚀状态进行检测,并取得了较好的结果。董明^[5]等人利用典型故障气体的相对含量在高维空间的分布特性输入到SVM(支持向量机)中对变压器故障类型诊断。薛延刚^[6]等人改进了SVM模型并引入故障分类准确性判定因子对水电机组进行了智能诊断。张丽平^[7]研究了PSO(粒子群)算法的基本结构、算法特点、改进方法、实现模式及应用等方面并将其引入到BP神经网络中,取得了满意的结果。蔡振宇^[8]等人将PSO-SVM(粒子群优化支持向量机)模型应用到振动机械故障诊断实例中,其结果表明相较于传统神经网络,PSO-SVM具有更高的准确性。

随机森林是通过集成学习的思想将多棵树(决策树)集成的一种算法,有着较高的准确率、可解释性及能够评估各个特征重要性等优势。Wang^[9]等人通过提取振动信号的无量纲指数做为特征参数,利用随机森林训练,预测精度高达100%。马富齐^[10]利用随机森林的特性通过剔除掉冗余的特征向量,进行了对机组的故障诊断。单增海^[11]等人将得到的多传感器信息融合后,进行特征筛选,利用多粒度级联森林模型完成了对液泵健康状态评估。段智勇^[12]等人利用随机森林算法对屏蔽泵进行故障诊断,并取得了满意的效果。徐书凡^[13]将潜油螺杆泵井的参数数据输入随机森林模型,完成了对潜油螺杆泵采油系统故障诊断。

当前在针对离心泵转子故障能够评估其故障特征重要性特征的研究较少,并且非重要特征之间的相关性会进一步削弱故障识别的效果,从而影响故障诊断效率及准确率。本文旨在通过随机森林对提取出的故障特征重要性进行分析、筛选,结合PSO-SVM对离心泵转子常见故障状态进行识别。通过在离心泵上放置振动加速度传感器进行原始信号采集,利用时域、频域处理方法,提取原始信号的时、频域特征参数矩阵,通过随机森林得到重要性较高的特征,并将之与随机森林分类得到的分类结果进行组合,得到新的故障特征参数矩阵,进入PSO-SVM中进行训练、验证。这样既能降低非重要故障特征对离心泵转子故障识别效果的影响,也能减少故障特征参数矩阵的维度,从而减少冗余的故障信息,降低计算的复杂程度。

1 随机森林计算特征重要度原理

随机森林(Random Forest,简称RF)是Bagging的一个扩展变体^[14]。它首先基于bagging思想,无放回的可重复的对初始数据集进行采样,再利用这些采样集训练决策树,同传统决策

树选择特征(信息增益)不同的是RF选择特征时随机从特征集中抽取一部分特征子集,并从这些子集中选择一个最优属性用于构建决策树的节点,进行下一步的分裂。由于构建每一颗决策树时是从数据集中进行随机可重复进行采样,因此还有部分数据集是没有被采样到的,这部分数据称为袋外数据(*oob*),将这部分数据代入已建立的决策树中,可以用于计算决策树的预测错误率(袋外数据误差,*oobErr*)。

由于原始信号具有信息量大、特征隐蔽和包含较多的噪声等特点,单纯直观分析无法获得被监测对象的具体健康状态信息,因此需要通过不同的信号处理方法对原始信号进行转换和处理,从而提取出能够反映运行设备的状态特征信息^[15]。均方根值、峰值、峰值因子、峭度、脉冲因子、裕度因子和波形因子是振动信号进行时域统计分析最常用的参数指标^[16]。为了更多反应原始信号携带的特征信息。另外选取了较为常用的4个时域特征和3个频域统计指标,时域特征为方差、均值、峭度因子、偏度;频域为重心频率、均方根频率、标准差频率。本文统计共14个时、频域指标作为故障特征的统计指标,计算公式如表1。

表中的 $x(n)$ 为信号的时域序列, $n=1,2,3,\dots,N$, N 为样本点数; $U(i)$ 表示的是信号 $x(n)$ 的频谱,其中 $i=1,2,3,\dots,m$, m 为谱线的个数; f_i 表示的是第 i 根谱线的频率值。

对于随机森林中某个特征重要性 a 的计算步骤如下:

(1) 假设随机森林中共有 n 颗决策树,对于一棵树 T_i ,其中 $i=1,2,3,\dots,n$,用袋外数据 $oob(i)$ 代入已建立的决策树 T_i 中,计算该树的袋外数据误差,记作 $oobErr01(i)$ 。

(2) 然后随机置换 $oob(i)$ 中第 j 列的特征参数,置换的方法是通过permutation方式将 $oob(i)$ 中所有样本的第 j 个特征参数重新打乱分布,保持其他特征参数不变,重新计算该树的袋外数据误差,记作 $oobErr02(i)$ 。

(3) 则该特征重要性 a 的计算公式如下:

$$a = \frac{oobErr02(i) - oobErr01(i)}{n} \quad (1)$$

若加入噪声后袋外数据准确率下降,即 $oobErr02(i)$ 上升,特征重要性 a 增大,则该特征重要程度比较高。

本文利用随机森林算法和PSO-SVM进行离心泵转子不平衡、不对中故障诊断的流程图见图1。

2 离心泵转子不平衡、平行不对中故障实验

以一台离心泵作为实验对象,离心泵的主要参数为:额定流量10.6 m³/h;额定转速2 900 r/min;额定扬程26 m;比转速49.88。信号采集选用的振动加速度传感器精度为 $\pm 1\%$ 。

振动加速度传感器布置在离心泵进口法兰水平径向。实验过程中先调节变频电机的转速使其达到设定转速,然后调节泵的出口阀门,使其达到设定的流量位置,待运行稳定时采集数据。

转子不平衡故障复现时分别用2.6、6.3、9 g的配重块安装在联轴器的螺栓上,使其产生质量偏心。不同重量的配重实验时,配重块安装在同一位置。为了提高该诊断模型在不同工况下的诊断率,根据泵的相似定律,每个配重实验分别在 $0.7n$ 、 $0.85n$ 、 $1.0n$ 转速工况, $0.7Q_n$ 、 $0.85Q_n$ 、 $1.0Q_n$ 流量工况下进行,共

表 1 特征计算方法

Tab.1 Calculation method of characteristics

名称	计算公式	名称	计算公式	名称	计算公式
均方根值	$\sqrt{\frac{\sum_{i=1}^N x_i^2}{N}}$	裕度因子	$\frac{\max(x(n))}{\left(\frac{1}{N} \sum_{n=1}^N \sqrt{ x(n) }\right)^2}$	偏度	$\frac{\max(x(n))}{\frac{1}{N} \sqrt{\sum_{n=1}^N x(n) ^2}}$
峰值	$\max(x(n))$	波形因子	$\frac{N \sqrt{\frac{1}{N} \sum_{n=1}^N (x(n))^2}}{\sum_{n=1}^N x(n)}$	重心频率	$\frac{\sum_{i=1}^m f_i U(i)}{\sum_{i=1}^m U(i)}$
峰值因子	$\frac{\max(x(n))}{\sqrt{\frac{1}{N} \sum_{n=1}^N (x(n))^2}}$	方差	$\frac{\sum_{n=1}^N (x(n) - \frac{\sum_{n=1}^N x(n)}{N})^2}{N}$	均方根频率	$\sqrt{\frac{\sum_{i=1}^m (f_i)^2 U(i)}{\sum_{i=1}^m U(i)}}$
峭度	$\sum_{n=1}^N (x(n) - \frac{\sum_{n=1}^N x(n)}{N})^4$	均值	$\frac{\sum_{n=1}^N x(n)}{N}$	标准差频率	$\sqrt{\frac{\sum_{i=1}^m (f_i - \frac{\sum_{i=1}^m f_i U(i)}{\sum_{i=1}^m U(i)})^2 U(i)}{m}}$
脉冲因子	$\frac{\max(x(n))}{\frac{1}{N} \sum_{n=1}^N x(n) }$	峭度因子	$\frac{\sum_{n=1}^N (x(n) - \frac{\sum_{n=1}^N x(n)}{N})^4}{N \left(\sqrt{\frac{1}{N} \sum_{n=1}^N (x(n))^2} \right)^4}$		

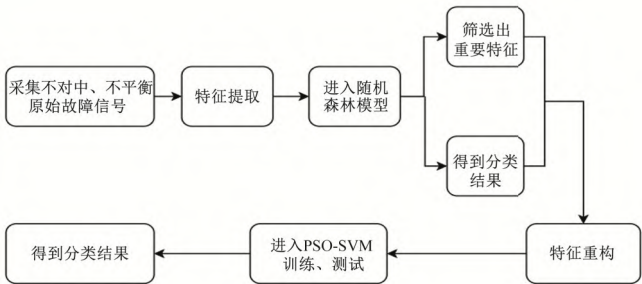


图 1 诊断流程图

Fig.1 Diagnostic flow chart



图 2 实验现场测点布置图

Fig.2 Layout of experimental points

27组实验。

转子平行不对中故障复现时,通过移动电机的位置,使转子产生不对中现象。不同平行不对中故障(30丝、40丝、50丝)实验时,分别在 $0.7n$ 、 $0.85n$ 、 $1.0n$ 转速工况, $0.7Q_n$ 、 $0.85Q_n$ 、 $1.0Q_n$

流量工况下进行,共27组实验。

3 实验数据分析与处理

实验所用传感器采集设定采集频率为25 600 Hz,采样时间1 s。其中正常、不平衡、不对中在 $1.0n$ 、 $1.0Q_n$ 工况下进口法兰水平径向的振动加速度信号时域图形如图3所示。

由于不平衡、不对中的特征频率主要出现在低频段,本文分析频谱图时只截取0~500 Hz作为分析频段。

图4为正常工况下、不平衡程度为6.3 g、不对中程度为30丝时额定转速、流量的轴承座测点的频域图。不平衡工况下对比正常工况下的频域图,可以看到图中一倍频(48.33 Hz)的幅值有所增大,这符合不平衡故障发生时的特征表现,频谱图中出现6倍频是由于实验泵所用叶轮为6叶片叶轮,出现的6倍频为叶频。不对中工况下对比相较于正常工况下的频域图,可以看到频谱图中二倍频的幅值有所增大,同时一倍频幅值有所降低,这符合转子平行不对中故障发生时的特征表现。

4 模型训练与分析

将实验采集获得的数据进行处理,按照上节特征提取方法提取故障特征参数,得到一个 315×14 的故障特征矩阵并进行标签标记,采用模型训练方法主要分为以下几个步骤进行。

(1) 故障特征集再构建,将故障特征集作为随机森林的输入,利用随机森林每次会有约1/3的样本没有参与决策树的建立,得到每个特征的重要性测度指标,将重要性前6个的故障特征及随机森林分类器得到的分类结果作为新特征集的组成,得

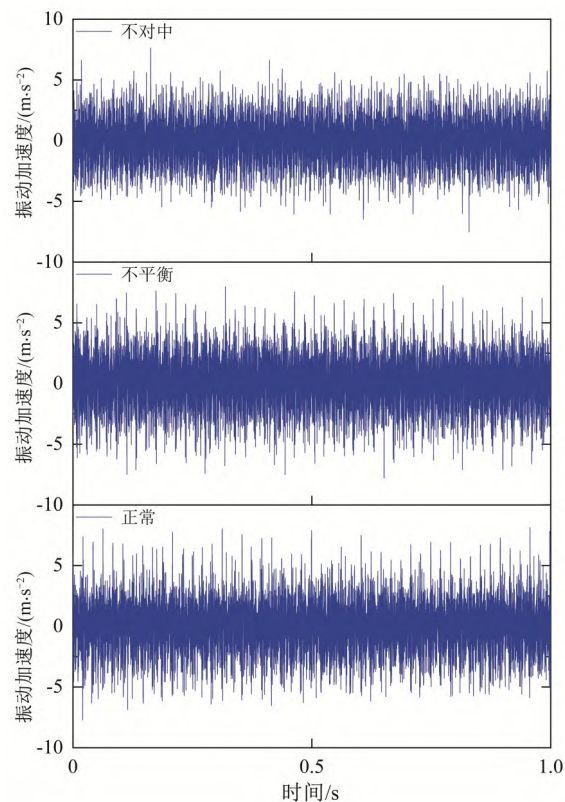


图3 不同设备状态下时域图

Fig.3 Time domain diagram under different device states

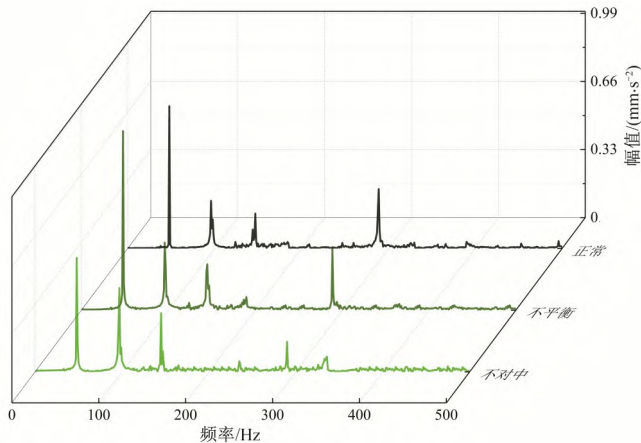


图4 频域图

Fig.4 Frequency domain diagram

到新故障特征集。

(2) SVM 算法属于有监督学习算法,是最优秀的分类算法之一,广泛的应用于科学技术领域,因此本文选择 SVM 为故障识别的学习算法。由于支持向量机(SVM)中核函数半径(g)、惩罚因子(c)难以选择最优的^[17],本文选择 RBF 作为 SVM 的核函数,利用粒子群优化支持向量机(PSO-SVM),寻找最优的粒子点为 SVM 的最优核函数半径和惩罚因子。将得到的新故障特征集等比例分组,50%作为 SVM 的训练集,用于训练模型;50%作为 SVM 的测试集,用于验证模型。

本文中,随机森林决策树数量设为 400,最小叶节点设置为 5,将原始特征矩阵输入得到每个特征的重要性评估由大到小排序如图 5 所示。

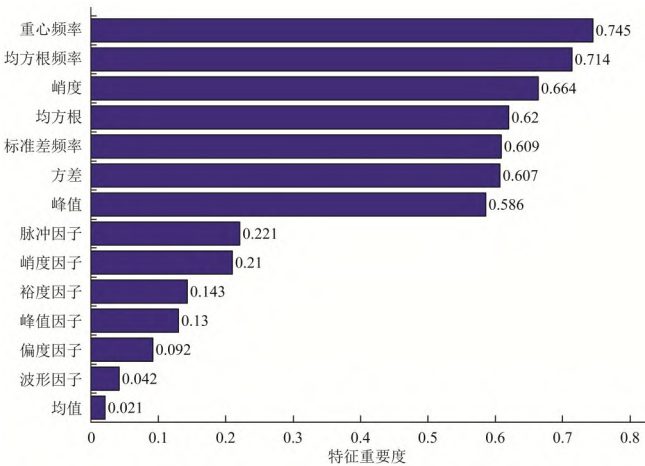


图5 特征重要性评估

Fig.5 Characteristics importance evaluation

取重要性排序前6的特征分别为峭度、方差、均方根、重心频率、均方根频率、标准差频率。同时将随机森林分类器得到的不同故障的分类概率结果作为新的故障特征,同重要性较高的6个特征最终得到一个新的315×9特征集。

4.1 支持向量机诊断的结果对比

故障特征矩阵进行分组,分组情况如表 2 所示。

表2 故障特征分组情况

Tab.2 Grouping of fault features

数据类型	训练集	标签结果	测试集	标签
正常工况	30	0	30	0
不平衡故障	90	1	90	1
不对中故障	90	2	90	2

将分组后的原始故障特征矩阵与新故障特征矩阵的训练集分别输入 PSO-SVM 中进行分类训练,PSO-SVM 适应度曲线如图 6 所示,从图 6 中可以看出重构后的原始特征在通过 PSO 算法计算最佳的适应度在经过第 4 次迭代后达到最优,此时的最优核函数半径 g 等于 0.01,惩罚因子 c 等于 12.395 9,SVM 训练精度 100%,相较于原始特征矩阵输入 SVM 中通过 PSO 算法计算最佳的适应度在经过第 33 次迭代后达到最优,训练精度为 85.6%,无论是迭代次数还是训练精度都有大幅度改善。

利用测试数据进行故障诊断,得到的结果如图 7。从图 7 中可以看出利用原始故障特征集进行验证时有大部分正常状态下的数据被识别成了不对中或不平衡,模型对正常状态下的故障特征数据识别表现不好,不能很好的区分正常与不对中故障。数据显示重构特征矩阵训练的 SVM 故障诊断模型对测试集的分类精度达到 99.36%,相较于原始故障训练的 SVM 故障诊断模型的识别精确度 86.7%,对故障的识别精度有明显的提升,其部分诊断概率详情见表 3。

5 结 论

利用随机森林筛选出的重要特征,并将其与随机森林的分类结果重新组合形成新的故障特征矩阵,通过将原始特征矩阵与重构后的特征矩阵分别输入进 PSO-SVM,得到每组特征向量对应不同标签的预测结果,主要的结论如下。

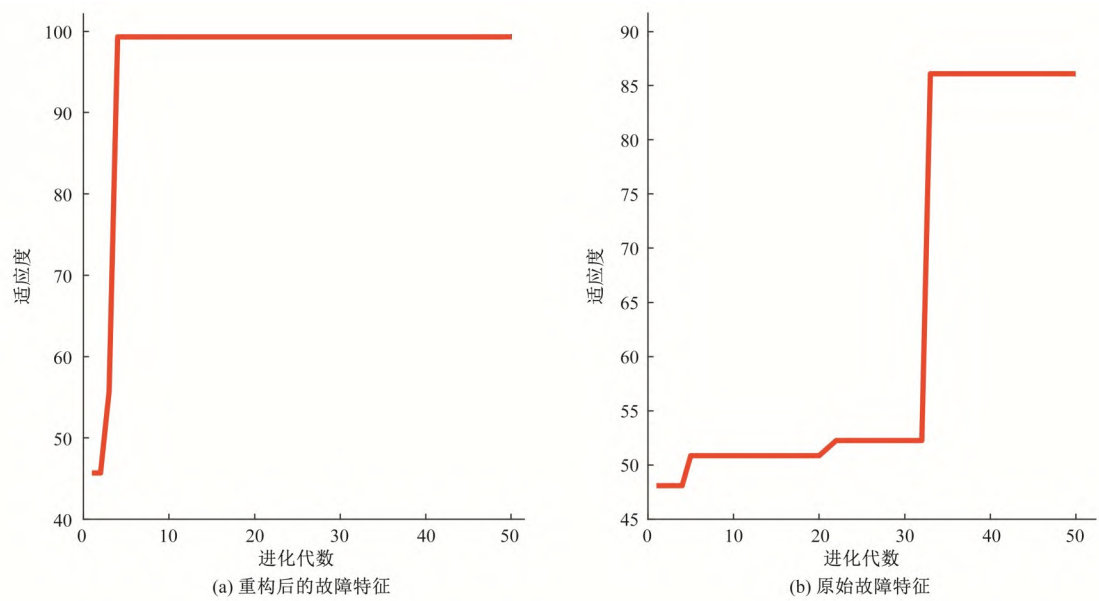


图6 PSO-SVM适应度曲线

Fig.6 PSO-SVM fitness curve

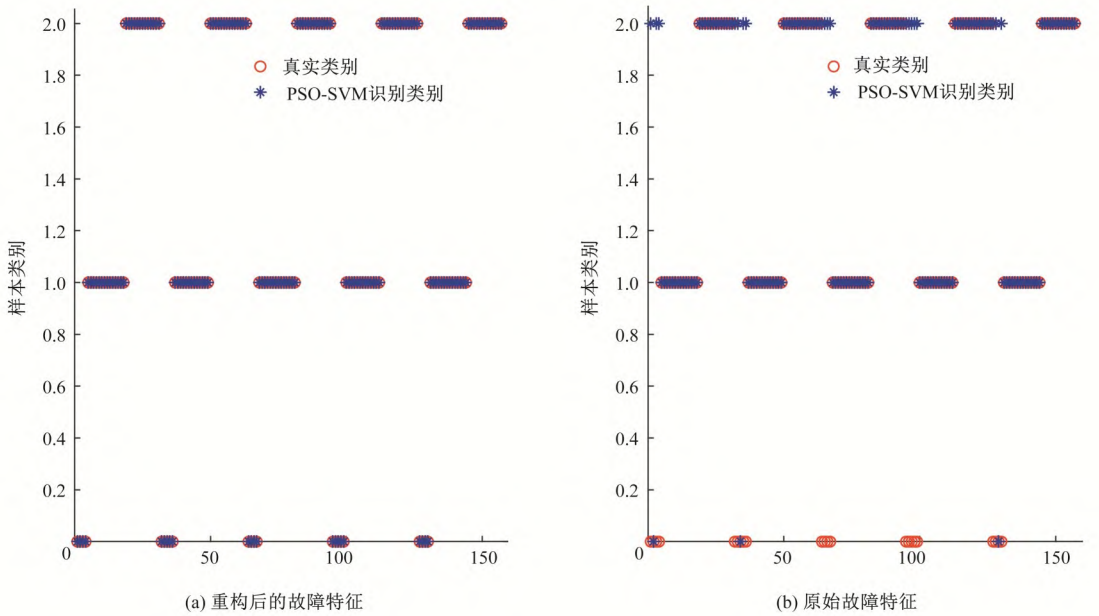


图7 进口法兰水平径向测点PSO-SVM诊断结果

Fig.7 PSO-SVM diagnosis results of horizontal and radial measurement points of inlet flange

表3 进口法兰水平径向测点部分概率详情

Tab.4 Partial probability details of inlet flange horizontal and radial measurement points

序号	标签结果为0的概率	标签结果为1的概率	标签结果为2的概率	预测结果	实际结果
1	0.925	0.034	0.040	0	0
2	0.893	0.020	0.087	0	0
3	0.918	0.022	0.061	0	0
...	...	...	...	...	...
12	0.008	0.977	0.015	1	1
13	0.008	0.977	0.015	1	1
...	...	...	...	...	...
208	0.047	0.007	0.946	2	2
209	0.011	0.012	0.977	2	2
210	0.007	0.005	0.988	2	2

(1) 重构后的故障特征矩阵在进入 PSO-SVM 中进行训练时,表现出更好的适应度,在寻优过程中仅迭代 4 次,便达到最佳训练精度,且训练精度明显高于原始故障矩阵的训练精度,达到 100%;

(2) 进行模型精度测试时,重构后的故障特征矩阵表现明显好于原始故障特征矩阵,其测试精度达到 99.36%,能够明显区分出正常、不对中、不平衡下的数据。

参考文献:

- [1] 曹冲锋. 基于 EMD 的机械振动分析与诊断方法研究[D]. 杭州: 浙江大学, 2009.
CAO C F. Study on mechanical vibration analysis and diagnosis methods based on empirical M-mode Decomposition [D]. Hangzhou: Zhejiang University, 2009.
- [2] 张彦铎, 姜兴渭. 多传感器信息融合及在智能故障诊断中的应用[J]. 传感器技术, 1999(3): 22-26.
ZHANG Y D, WEI X W. Multi-Sensor information fusion technique and its application on intelligent fault diagnosis system [J]. Chinese Journal of Sensors and Actuators, 1999(3): 22-26.
- [3] 屈梁生. 机械故障的全息诊断原理[M]. 北京: 科学出版社, 2007.
- [4] DONG L, WU K, ZHU J C, et al. Cavitation detection in centrifugal pump based on interior flow borne noise using WPD-PCA-RBF[J]. Shock and Vibration, 2019.
- [5] 董明, 孟源源, 徐长响, 等. 基于支持向量机及油中溶解气体分析的大型电力变压器故障诊断模型研究[J]. 中国电机工程学报, 2003(7): 88-92.
DONG M, MENG Y Y, XU C X, YAN Z. Fault diagnosis model for power transformer based on support vector machine and dissolved gas analysis [J]. Proceedings of the CSEE, 2003(7): 88-92.
- [6] 薛延刚, 王勇劲, 赵海英, 等. 基于时序贴近度与改进 SVM 的水机轴心轨迹诊断[J]. 排灌机械工程学报, 2017, 35(12): 1 054-1 057+1 062.
XUE Y G, WANG J Y, ZHAO H Y, LIU G W, QIAO F. Axis orbit automatic identification on time-series similarity mining for hydropower units [J]. Journal of Drainage and Irrigation Machinery Engineering, 2017, 35(12): 1 054-1 057+1 062.
- [7] 张丽平. 粒子群优化算法的理论及实践[D]. 杭州: 浙江大学, 2005.
ZHANG L P. The theorem and practice upon the particle swarm optimization algorithm [D]. Hangzhou: Zhejiang University, 2005.
- [8] 蔡振宇, 史乔波. 基于 PSO-SVM 模型的旋转机械故障诊断研究[J]. 煤矿机械, 2011, 32(11): 276-278.
CAI Z Y, SHI Q B. Fault Diagnosis of rotating machinery based on PSO-SVM Model [J]. Coal Mine Machinery, 2011, 32(11): 276-278.
- [9] WANG Z W, ZHANG Q H, XIONG J B, et al. Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests [J]. IEEE Sensors Journal, 2017, 17(17): 5 581-5 588.
- [10] 马富齐. 基于复数据经验模态分解和随机森林理论的风电机组多域特征故障诊断研究[D]. 西安: 西安理工大学, 2019.
MA F Q. Research on multi-domain feature fault diagnosis of wind turbine based on complex empirical mode decomposition and random forest theory [D]. Xi'an: Xi'an University of Technology, 2019.
- [11] 单增海, 李志远, 张旭, 等. 基于多传感器信息融合和多粒度级联森林模型的液压泵健康状态评估[J]. 中国机械工程, 2021, 32(19): 2 374-2 382.
SAN Z H, LI Z Y, ZHANG X, et al. Health status assessment of hydraulic pump based on multi-sensor information fusion and multi-grained cascade forest model [J]. China Mechanical Engineering, 2021, 32(19): 2 374-2 382.
- [12] 段智勇, 刘才学, 艾琼, 等. 基于随机森林的屏蔽泵故障诊断方法研究[J]. 核科学与工程, 2020, 40(4): 625-630.
DUAN Z Y, LIU C X, AI Q, HE P. Research on fault diagnosis method of canned motor pump based on random forests [J]. Nuclear Science and Engineering, 2020, 40(4): 625-630.
- [13] 徐书凡. 电动潜油螺杆泵系统故障诊断研究[D]. 黑龙江大庆: 东北石油大学, 2020.
XU S F. Fault diagnosis of electric submersible screw pump system [D]. Heilongjiang Northeast Petroleum University, 2020.
- [14] BREIMAN L. (2001a). Random Forest. Machine Learning, 45(1): 5-32.
- [15] 吴侃. 离心泵空化状态识别方法研究与实现[D]. 江苏镇江: 江苏大学, 2019.
WU K. Research and implementation of cavitation state recognition method for centrifugal pump [D]. Zhenjiang: Jiangsu University, 2019.
- [16] 万书亭, 吴美玲. 基于时域参数趋势分析的滚动轴承故障诊断[J]. 机械工程与自动化, 2010(3): 108-110+113.
WAN S T, WU M L. Fault diagnosis for rolling bearing based on trend analysis of time domain parameters [J]. Mechanical Engineering & Automation, 2010(3): 108-110+113.
- [17] 平源. 基于支持向量机的聚类及文本分类研究[D]. 北京: 北京邮电大学, 2012.
PING Y. Research on clustering and text categorization based on support vector machine [D]. Beijing: Beijing University of Posts and Telecommunications, 2012.

欢迎投稿 欢迎订阅 欢迎刊登广告